CORE DISCUSSION PAPER

2003/XX

SEMIPARAMETRIC MULTIVARIATE GARCH MODELS

C.M.Hafner[1] , *J.V.K.Rombouts*[2]

January 2003

Abstract

Estimation of multivariate GARCH models is usually carried out by quasi maximum likelihood (QMLE), for which recently consistency and asymptotic normality have been proven under quite general conditions. However, there are to date no results on the efficiency loss of QMLE if the true innovation distribution is not multinormal. We investigate this issue by suggesting a nonparametric estimation of the multivariate innovation distribution, based on consistent parameter estimates obtained by QMLE. We give conditions under which the semiparametric efficiency bound can be attained. A simulation experiment demonstrates the efficiency gain of our procedure compared with QMLE, and an application to a bivariate stock index series illustrates the results.

Keywords: Multivariate GARCH models, semiparametric methods, efficient estimation.

JEL Classification: C14, C22.

[1]Econometric Institute, Erasmus University Rotterdam, P.O.B. 1738, 3000 DR Rotterdam, Netherlands, `chafner@few.eur.nl`

[2]CORE and Department of Economics, Université catholique de Louvain, Voie du Roman Pays 34, 1348, Louvain-la-Neuve, Belgium, `rombouts@core.ucl.ac.be`.

The authors would like to thank Luc Bauwens, Geert Dhaene, Oliver Linton and Bas Werker for helpful discussions. We also thank participants of the CORE econometrics seminar, the York annual meeting in econometrics, and the 2002 annual econometric study group in Bristol.

This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

1 Introduction

When modelling comovements in financial time series multivariate GARCH models are often used. Broadly speaking, these models are data driven in the sense that when conditioning on an information set the conditional variance covariance matrix of the process is known. In principle, estimation of the models is straightforward if one supposes a specific parametric distribution of the innovations. The model then becomes purely parametric and is presumed to be known completely up to some parameters that are estimated by the maximum likelihood method. Quite problematic is the choice of the distribution of the innovations. Usually one assumes normality, which provides the so-called quasi maximum likelihood estimates (QMLE). It is now well-known that QMLE is consistent even if the true underlying distribution is not normal, see e.g. Bollerslev and Wooldridge (1992) and Jeantheau (1998). However, in the case of misspecification there may be a substantial efficiency loss of QMLE compared with the correctly specified maximum likelihood estimator (MLE). On the other hand, assuming a non-normal distribution entails the risk of inconsistent parameter estimation if the distribution is misspecified.

In this paper we propose an alternative approach. We still assume that the variance covariance matrix has a specific parameterized structure. However, we follow a nonparametric approach in letting the data determine the distribution of the innovations. This semiparametric (SP) approach results in models that are in a more general class than the fully parametric class of models. With the typically large data sets in finance we would expect to obtain density estimates that are reasonably close to the true distribution of the innovations. In a univariate framework, semiparametric GARCH models have been proposed and discussed by Engle and Gonzalez-Rivera (1991), Linton (1993), and Drost and Klaassen (1997). Whether or not it is possible to obtain efficient semiparametric estimates in the multivariate case is one of the topics we investigate. It turns out that estimators can be constructed that achieve the semiparametric lower bound and, hence, are efficient, if the true distribution is known to belong to the class of spherical distributions. For this case, we provide conditions under which the estimators would even be adaptive (and hence achieve the parametric lower bound). We show that the normal distribution is the only distribution that fulfills these conditions, so that adaptive estimation is not possible in general. We characterize several spherical distributions with respect to their distance from the parametric lower bound. In the general case of non-spherical distributions it seems very difficult if not impossible to construct SP-efficient estimators. In a simulation study using leptokurtic distributions, we show that there are substantial efficiency gains of SP over QMLE.

The paper is organized as follows. First, the model framework and the traditional estimation method is introduced. The third section discusses the nonparametric estimation of the innovation distribution, as well as the efficiency of SP estimators. In the fourth section, a simulation study is provided, and the last section presents results of an application to two stock index series, the

CAC40 and the FTSE100. The proofs of the propositions are given in Appendix D.

2 The multivariate GARCH Model

Consider $\{y_t\}$ a vector stochastic process of dimension N with a countable index set and an uncountable state space. We write

$$y_t = \mu_t + \epsilon_t \tag{1}$$

where μ_t is a vector of dimension N that may depend on past information up to time $t-1$. In this paper the conditional mean μ_t is not of particualar interest, so we assume in the following that $\mu_t = 0$. The error vector ϵ_t is factorized as

$$\epsilon_t = H_t^{1/2}(\theta)v_t,$$

where $H_t(\theta)$ is a positive definite matrix that also may depend on past information up to time $t-1$, and θ is a finite dimensional parameter vector, $\theta \in \Theta \subset R^K$. Hereafter, we will suppress the dependence of H_t on θ for notational convenience.

Assumption 1 *$\{v_t\}$ is an i.i.d. sequence with $E[v_t] = 0$ and $E[v_t v_t'] = I_N$ where I_N is the identity matrix of dimension N. The function $g(\eta)$ is the density function of v_t which is indexed by a possibly infinite dimensional vector η. The vector η can be regarded as a nuisance parameter in our framework.*

Clearly, the conditional mean of y_t is just given by the vector μ_t and the conditional variance matrix by H_t. As usual, we condition on the sigma field generated by all the information (here the y_t's) until time $t-1$. The set $\mathcal{F}_{t-1}$ contains all this information.

Several specifications are proposed for H_t. See Bauwens, Laurent and Rombouts (2002) for a survey. The so–called vec representation of a multivariate GARCH(p,q) model is given as

$$h_t = \mathrm{vech}(H_t) = \omega + \sum_{i=1}^{q} A_i \mathrm{vech}(\varepsilon_{t-i}\varepsilon_{t-i}') + \sum_{i=1}^{p} B_i \mathrm{vech}(H_{t-i}), \tag{2}$$

where A_i and B_i are $N^* \times N^*$ parameter matrices, and ω is an N^* parameter vector with $N^* = N(N+1)/2$.

Engle and Kroner (1995) discuss in detail a dynamic specification of the following form:

$$H_t = C_0 C_0' + \sum_{k=1}^{K}\sum_{i=1}^{q} A_{ki}'\varepsilon_{t-i}\varepsilon_{t-i}' A_{ki} + \sum_{k=1}^{K}\sum_{i=1}^{p} B_{ki}' H_{t-i} B_{ki}, \tag{3}$$

where C_0 is a lower triangular matrix and A_{ki} and B_{ki} are $N \times N$ parameter matrices. By applying the vech operator to both sides of (3), it is easily seen that the BEKK specification is a special case of the vec model. The link between the parameters is given by $\omega = \mathrm{vech}(C_0 C_0')$,

$A_i = \sum_{k=1}^{K} D_N^+ (A_{ki} \otimes A_{ki})' D_N$ and $B_i = \sum_{k=1}^{K} D_N^+ (B_{ki} \otimes B_{ki})' D_N$, where D_N is the duplication matrix defined in (66) and D_N^+ is its generalized inverse, see (62). However, compared with the more general model the BEKK specification entails several practical advantages. For example, it will often be more parsimonious. Furthermore, for given positive definite initial covariances $H_0, \ldots, H_{1-p}$, the BEKK representation generates sample covariances $H_t, t = 1, \ldots, n$, that are positive definite under the weak (sufficient) condition that at least one of the matrices C_0 or B_{ki} has full rank (Engle and Kroner 1995).

Assumption 2 *The eigenvalues of* $\sum_{i=1}^{q} A_i + \sum_{j=1}^{p} B_j$ *have modulus smaller than one.*

This is a necessary and sufficient condition for the existence of the unconditional variance matrix of ϵ_t. The matrix is then given by $\Sigma = V(\varepsilon_t)$ with

$$\mathrm{vech}(\Sigma) = \left(I_{N^*} - \sum_{i=1}^{q} A_i - \sum_{j=1}^{p} B_j \right)^{-1} \omega.$$

If one supposes that the distribution of v_t is known then maximum likelihood estimation (MLE) is in principle straightforward. Nevertheless, because the number of parameters is often large, estimation can become a tedious exercise. If v_t is assumed to be normally distributed with zero mean vector and I_N variance matrix then ϵ_t will be conditionally normally distributed with zero mean vector and H_t as covariance matrix. The likelihood for a sample of n observations then takes the form

$$L^{qml}(\theta) = -\frac{Nn}{2} \log(2\pi) - \sum_{t=1}^{n} \frac{1}{2} \log |H_t| - \frac{1}{2} \varepsilon_t' H_t^{-1} \varepsilon_t. \tag{4}$$

Defining

$$l_t^{qml}(\theta) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log |H_t| - \frac{1}{2} \varepsilon_t' H_t^{-1} \varepsilon_t,$$

we can write $L^{qml}(\theta) = \sum_{t=1}^{n} l_t^{qml}(\theta)$. As shown by Bollerslev and Wooldridge (1992) in a general conditional heteroskedasticity framework, maximizing (4) provides consistent estimates even if the likelihood is misspecified under fairly general conditions. Therefore this method has been termed Quasi Maximum Likelihood (QML) estimation. Again under general conditions, the asymptotic distribution of QML parameter estimates $\tilde{\theta}$ is given by

$$\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{\mathcal{D}} N(0, V_{qml})$$

with $V_{qml} = \mathcal{J}^{-1} \mathcal{I} \mathcal{J}^{-1}$, and

$$\mathcal{J} = -\mathrm{E}\left[\frac{\partial^2 l_t^{qml}}{\partial \theta \partial \theta'} \right], \quad \mathcal{I} = \mathrm{E}\left[\frac{\partial l_t^{qml}}{\partial \theta} \frac{\partial l_t^{qml}}{\partial \theta'} \right],$$

where expectations are taken with respect to the true distribution and are evaluated at the true parameter vector θ. In the case of correct specification, i.e., the distribution of v_t is indeed multinormal, $\mathcal{J} = \mathcal{I}$, and $V_{qml} = \mathcal{I}^{-1}$.

While QML estimates are often consistent, they are inefficient if the likelihood is misspecified. Therefore one sometimes considers the multivariate t distribution as an appropriate choice because of potential fat tails in the innovations. The drawback of this assumption is that if the assumption of a specific non-normal distribution is not correct, then in general the parameter estimates are not even consistent, as shown by e.g. Bollerslev and Wooldridge (1992). Therefore, we will not pursue the assumption of a specific parametric distribution in our paper. In the next section we formalize our motivation for giving all the weight to the data in search for a suitable distribution.

3 Semiparametric GARCH estimation

This section describes the methodology used to obtain semiparametric GARCH estimates. A detailed description of semiparametric estimation techniques and adaptivity is beyond the scope of this paper. We refer to Bickel (1982), Newey (1990) and Drost, Klaassen and Werker (1997) for details.

We consider two approaches. One approach is quite similar to the one employed by Engle and Gonzalez-Rivera (1991) for univariate GARCH models. For practical reasons, this methodology may be preferred, but efficiency gains might still be possible. Another approach is the one used by Drost and Klaassen (1997) to obtain efficient estimates. It requires to modify the estimated score by a projection on the so-called tangent set, which may not always be possible in practice. This procedure is more ambitious but also somewhat more difficult to implement. It will be discussed in the next section.

To obtain estimates of the parameter vector θ for a given sample of n observations, we maximize the log likelihood

$$L(\theta) = -1/2 \sum_{t=1}^{n} \log |H_t| + \sum_{t=1}^{n} \log g(H_t^{-1/2} \varepsilon_t), \tag{5}$$

where g is an unspecified density function of standardized residuals. Note that Assumption 1 states that g is a density with mean zero and identity covariance matrix. Without this assumption the model would not be identified. In the first approach, the idea to estimate g is to first use QMLE (i.e. Gaussian g) to obtain standardized residuals, and then estimate nonparametrically the density g. The steps are suggested in Algorithm 1:

Algorithm 1

1. *Use QMLE to obtain a consistent estimate of θ, $\tilde{\theta}$, say, that gives $\tilde{H}_t$.*

2. *Calculate standardized residuals, $\tilde{v}_t = \tilde{H}_t^{-1/2} \varepsilon_t$. Make sure that they have mean zero and variance I_N.*

3. *Estimate nonparametrically the density g of $\tilde{v}_t$, giving $\hat{g}$.*

4. Maximize L keeping $\hat{g}$ fixed.

This procedure can be viewed as a generalization of the one suggested for univariate GARCH models by Engle and Gonzalez-Rivera (1991).

For the nonparametric density estimation, we use kernel estimates. A general multivariate kernel density estimator with bandwidth matrix H and multivariate kernel $\mathcal{K}$ can be written as

$$\hat{g}_H(x) = \frac{1}{n|H|}\sum_{t=1}^{n}\mathcal{K}(H^{-1}(v_t - x))$$

Since the scale of the variables should be the same (same variance in all directions), it is reasonable to use a scalar bandwidth, $H = hI_N$, with $h > 0$. It is well known that by requiring $nh^N \to \infty$ and $h \to 0$ as $n \to \infty$, the multivariate kernel density estimates are consistent and asymptotically normally distributed. The MSE-optimal rate for the bandwidth is $n^{-1/(4+N)}$. We use here a rule of thumb bandwidth as proposed by Silverman (1986). Furthermore, we use a product kernel $\mathcal{K}(x) = \prod_{i=1}^{N} K(x_i)$ and some univariate kernel function K such as Gaussian, quartic or Epanechnikov. Thus, our density estimate becomes

$$\hat{g}_h(x) = \frac{1}{nh^N}\sum_{t=1}^{n}\prod_{i=1}^{N} K(\frac{v_{i,t} - x_i}{h})$$

For details on multivariate kernel density estimation see the excellent survey of Scott (1992).

Alternatively, one can estimate the semiparametric model using a two-step procedure that uses the so-called influence function to correct an initial consistent QMLE estimate. The correction is essentially a one-step Newton-Raphson algorithm based on the score vector of the likelihood. Asymptotically, this algorithm is equivalent to the previously discussed iterative procedure. Formally, let us write the log likelihood as $L(\theta) = \sum_{t=1}^{n} l_t(\theta)$ with

$$l_t(\theta) = -\frac{1}{2}\log|H_t| + \log g(H_t^{-1/2}\varepsilon_t).$$

The next proposition provides a formula for the score vector.

Proposition 1 *The score vector takes the form*

$$\dot{l}_t(\theta) \equiv \frac{\partial l_t(\theta)}{\partial\theta} = \frac{\partial vec(H_t)'}{\partial\theta}\left\{-\frac{1}{2}vec(H_t^{-1}) - D_N D_N^{+}(I_N \otimes H_t + I_{N^2})^{-1} vec\left(\frac{\partial \log g(v_t)}{\partial v_t} v_t'\right)\right\}. \tag{6}$$

A semiparametric estimate of θ is given by

$$\hat{\theta} = \tilde{\theta} + \left(\sum_{t=1}^{n}\dot{\ell}_t(\tilde{\theta})\dot{\ell}_t(\tilde{\theta})'\right)^{-1}\sum_{t=1}^{n}\dot{\ell}_t(\tilde{\theta}) \tag{7}$$

where $\tilde{\theta}$ is the initial QMLE estimate and $\dot{\ell}_t(\tilde{\theta})$ is estimated using the nonparametric estimate of g ($\hat{g}$). We may expect this estimate to work asymptotically as well as the iterative estimate, but its computational burden is much weaker. However, as already noticed by Engle and Gonzalez-Rivera (1991), this semiparametric estimate is not likely to achieve the semiparametric lower bound in general. As we will see in the next section, one can do better if the distribution g is known (or cannot be rejected) to belong to the class of spherical distributions.

3.1 Efficient semiparametric estimation

In the special case where $\dot{\ell}_t$ can be factorized into a term that only depends on history and another term that only depends on the density, one can find semiparametric estimators that achieve the semiparametric lower bound and are therefore efficient. Such a factorization does not seem to be possible for the general score function in (6) unless one makes a symmetry assumption about the matrix $\frac{\partial g(v_t)}{\partial v_t}v_t'$. Written elementwise, this means that $\forall i, j$,

$$\frac{\partial g(x)}{\partial x_i}x_j = \frac{\partial g(x)}{\partial x_j}x_i, \tag{8}$$

which is a system of linear partial differential equations. The solutions are of the form $g(x) = f(x_1^2 + \ldots + x_N^2)$, but this is just the definition of the so-called spherical distributions[1].

Assumption 3 *The innovation density g belongs to the class of spherical distributions.*

Examples of spherical distributions are the multivariate versions of the normal, the t, the logistic and Laplace distributions.

Proposition 2 *Under Assumption 3, the score vector reduces to*

$$\dot{l}_t(\theta) = W_t(\theta)D_N\psi_t.$$

with the $(K \times N^2)$ matrix $W_t = (vec(W_{t1}), \ldots, vec(W_{tK}))'$, where $W_{ti} = \frac{1}{2}H_t^{-1/2}\frac{\partial H_t}{\partial \theta_i}H_t^{-1/2}$, and the $(N^* \times 1)$ vector $\psi_t = -vech\left(I_N + \frac{\partial \log g(v_t)}{\partial v_t}v_t'\right)$.

Note that W_{ti} depends only on past information and on the specification for H_t. The other term ψ_t depends only on the innovation v_t, so that W_{ti} and ψ_t are stochastically independent. An important result is that under Assumption 3, the score vector is a martingale difference sequence, i.e.,

$$\mathrm{E}\left[\dot{l}_t(\theta) \mid \mathcal{F}_{t-1}\right] = 0,$$

which is typically used for deriving the asymptotic distribution of the parameter estimates. In the Gaussian case, denote the score vector by F_t, with $F_t = \mathrm{vech}(v_tv_t' - I_N)$, and let $M_{\psi\psi} = \mathrm{E}[\psi_t\psi_t']$, $M_{\psi F} = \mathrm{E}[\psi_t F_t']$, $M_{F\psi} = \mathrm{E}[F_t\psi_t']$, and $M_{FF} = \mathrm{E}[F_tF_t']$. Note that $M_{\psi F}' = M_{F\psi}$ and $M_{F\psi}' = M_{\psi F}$. To ensure that these matrices exist, we make the following assumption.

[1] Suppose X has a density $g(x) = g(x_1, \ldots, x_N)$. This is a spherical density when $g(x) = f(x'x)$ for some function $f : R^+ \to R^+$.

Assumption 4 *All fourth order moments of v_t exist, and*

$$E\left[\left(\frac{\partial \log g(x)}{\partial(x'x)}\right)^2\right] < \infty$$

Assumption 4 excludes, for example, a multivariate Student t distribution with 4 or less degrees of freedom.

The efficient semiparametric estimate is now given by

$$\hat{\theta} = \tilde{\theta} + \left(\sum_{t=1}^{n} \dot{\ell}_t^*(\tilde{\theta})\dot{\ell}_t^{*'}(\tilde{\theta})\right)^{-1} \sum_{t=1}^{n} \dot{\ell}_t^*(\tilde{\theta}) \tag{9}$$

where

$$\dot{\ell}_t^*(\theta) \quad = \quad \dot{\ell}_t(\theta) - P_t \tag{10}$$

and

$$P_t = E[W_t(\theta)]D_N\{\psi_t - M_{\psi F}M_{FF}^{-1}F_t\}. \tag{11}$$

If we define $S(\eta)$ as the population score for the nuisance parameters then $\dot{\ell}_t^*(\tilde{\theta})$ is found by projection of $\dot{\ell}_t$ on the closure set of all linear combinations of $S(\eta)$, called the tangent set.

Under regularity conditions, the asymptotic distribution of the semiparametric estimator is given by

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N(0, V_{sp})$$

with $V_{sp} = \mathrm{E}[\dot{\ell}_t^*\dot{\ell}_t^{*\prime}]^{-1}$.

By definition, adaptive estimation is possible if and only if $P_t = 0$, which means that the semiparametric efficient score, $\dot{\ell}_t^*$, is equal to the parametric score $\dot{\ell}_t$ and, hence, V_{sp} is equal to the parametric lower bound, $V_{ml} = \mathrm{E}[\dot{\ell}_t\dot{\ell}_t']^{-1}$.

Proposition 3 *Under Assumptions 3 and 4,*

$$V_{ml}^{-1} \quad = \quad E[W_t D_N M_{\psi\psi} D_N' W_t'] \tag{12}$$

$$V_{sp}^{-1} \quad = \quad E[W_t D_N M_{\psi\psi} D_N' W_t'] - E[W_t] D_N Q D_N' E[W_t'] \tag{13}$$

$$V_{qml}^{-1} \quad = \quad E[W_t D_N M_{\psi F} D_N' W_t'] E[W_t D_N M_{FF} D_N' W_t']^{-1} E[W_t D_N M_{F\psi} D_N' W_t'] \tag{14}$$

with

$$Q = M_{\psi\psi} - M_{\psi F} M_{FF}^{-1} M_{F\psi} \tag{15}$$

As a corollary, we obtain that the difference of the information between the MLE and the semiparametric estimator is given by the positive semi-definite matrix

$$V_{ml}^{-1} - V_{sp}^{-1} = \mathrm{E}[W_t] D_N Q D_N' \mathrm{E}[W_t'].$$

The matrix Q determines the inefficiency of the SP estimator w.r.t. MLE. Adaptive estimation is possible if and only if $Q = 0$. Similar to Gonzalez-Rivera and Drost (1999) it can also be verified that $V_{sp}^{-1} - V_{qml}^{-1}$ is positive semi-definite.

The matrix M_{FF} depends on the structure of fourth moments of v_t. Lemma 5 implies that the marginal kurtosis $\kappa = E[v_{ti}^4]$ is linked to any co-kurtosis $c = E[v_{ti}^2 v_{tj}^2], j \neq i$ by $\kappa = 3c$, and M_{FF} depends on only one parameter.

Proposition 4 *Under Assumptions 3 and 4, we have*

$$\begin{aligned} M_{FF} &= 2cD_N^+ D_N^{+\prime} + (c-1) vech(I_N) vech(I_N)' && (16) \\ M_{\psi\psi} &= 2\tau D_N^+ D_N^{+\prime} + (\tau - 1) vech(I_N) vech(I_N)' && (17) \\ M_{\psi F} &= 2D_N^+ D_N^{+\prime} = M_{F\psi} && (18) \end{aligned}$$

where

$$\tau = E\left[\left(\frac{\partial \log g(x)}{\partial x_1} \right)^2 x_1^2 \right] /3. \qquad (19)$$

For example, in the bivariate case ($N = 2$) we have

$$M_{\psi F} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad M_{FF} = \begin{pmatrix} 3c-1 & 0 & c-1 \\ 0 & c & 0 \\ c-1 & 0 & 3c-1 \end{pmatrix},$$

and

$$M_{\psi F} M_{FF}^{-1} M_{F\psi} = \frac{1}{c(2c-1)} \begin{pmatrix} 3c-1 & 0 & 1-c \\ 0 & 2c-1 & 0 \\ 1-c & 0 & 3c-1 \end{pmatrix}.$$

In practice, the moment matrices $M_{\psi F}$ and M_{FF} can be replaced by the empirical moments of the innovations $\hat{v}_t$ and score $\hat{\psi}_t$ obtained after the first step.

Clearly, adaptivity is possible in the Gaussian case since then by definition $\psi_t = F_t$. Whether there are other distributions in the spherical class that allow for adaptivity is our next concern. In a univariate framework, Gonzalez-Rivera (1997) has shown that a class of symmetric bimodal distributions allows for adaptivity, but when restricting the distributions to unimodality, the Gaussian distribution is the only distribution under which adaptivity is possible. The following proposition states that this extends to the multivariate case, since spherical distributions are symmetric and unimodal.

Proposition 5 *In the class of spherical distributions, the multinormal is the only one that allows for adaptivity.*

Let us now look at three examples of spherical distributions, the Student t, Laplace and logistic distribution. Table 1 reports the spectral norm of Q for these distributions, which for the case of real, positive semi-definite matrices is equal to the spectral radius $\rho(Q)$, i.e., the largest eigenvalue.

	$N=1$			$N=2$			$N=3$		
Distribution	c	τ	$\rho(Q)$	c	τ	$\rho(Q)$	c	τ	$\rho(Q)$
t ($\nu=5$)	3.0000	0.7500	0.7500	3.0000	0.7778	0.8889	3.0000	0.8000	0.9333
t ($\nu=8$)	1.5000	0.8182	0.3117	1.5000	0.8333	0.3333	1.5000	0.8461	0.3590
t ($\nu=12$)	1.2500	0.8667	0.1454	1.2500	0.8750	0.1667	1.2500	0.8824	0.1810
Laplace	2.0000	0.6667	0.2000	1.6667	0.7500	0.3000	1.5000	0.8000	0.2667
ES Logistic	0.7948	5.0965	11.4005	0.8559	1.4431	0.9629	0.8960	1.2968	0.7816

Table 1: Co-kurtosis c, the value of τ in (19), and the spectral norm of the matrix Q in (15)

1. The density of a symmetric standardized multivariate Student t distribution is given by

$$g(v_t) = \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\{\pi(\nu-2)\}^{N/2}\Gamma(\nu/2)}\left(1+\frac{v_t'v_t}{\nu-2}\right)^{-(\nu+N)/2} \tag{20}$$

   where $\Gamma(p) = \int_0^\infty x^{p-1}e^{-x}dx$ is the gamma function. To ensure finite fourth moments of v_t we will assume in the following that $\nu > 4$. Under the density given in (20), $c = (\nu-2)/(\nu-4)$ and $\tau = (\nu+N)/(\nu+N+2)$. Note that for $\nu \longrightarrow \infty$, since the limiting distribution is a Gaussian, $c = 1$, $\tau = 1$, and $M_{\psi\psi} = M_{\psi F} = M_{FF} = 2D_N^+D_N^{+\prime}$. For increasing dimensions N, τ converges to 1 and $M_{\psi\psi}$ converges to $2D_N^+D_N^{+\prime}$.

2. The second example is a multivariate Laplace distribution with density

$$g(v_t) = \frac{(N+1)^{N/2}\Gamma(N/2)}{2\pi^{N/2}(N-1)!}\exp(-\sqrt{(N+1)v_t'v_t}) \tag{21}$$

   For $N = 1, 2, 3$ we find $c = (N+3)/(N+1)$ and $\tau = (N+1)/(N+2)$. Although we do not use it here, we conjecture that these formulae for c and τ hold for any N, which would imply that $c \to 1$ and $\tau \to 1$ for $N \to \infty$, which in turn implies using Proposition 4 that the multivariate Laplace density converges to a multinormal distribution with increasing dimension.

3. The third example is an elliptically symmetric (ES) multivariate logistic distribution with density

$$g(v_t) = c_1 \frac{e^{-c_2 v_t'v_t}}{(1+e^{-c_2 v_t'v_t})^2} \tag{22}$$

   with constants c_1 and c_2 such that (22) integrates to one and $\mathrm{Var}(v_{it}) = 1$. We calculate the values c_1, c_2, the co-kurtosis C, and τ by numerical integration. Note that the univariate distribution ($N = 1$) is different from distribution usually called logistic. For example, the distribution in (22) is platykurtic, whereas the usually called logistic is leptokurtic. The reason for not considering the standard multivariate logistic is that it does not belong to the class of spherical distributions. The ES logistic distribution is mentioned by Jensen (1985).

Note that the results reported in Table 1 generalize those for the case $N = 1$ listed by Gonzalez-Rivera and Drost (1999), except for the logistic distribution which, as explained, is defined in a different way. In the univariate case, Q is a positive scalar, so that $\rho(Q)$ is just this scalar itself.

4 Monte Carlo simulation experiment

In this section we are interested in the performance of the proposed SP estimator relative to the QML and ML estimator. Intuitively, the semiparametric method should perform better than QML, but worse than ML, when there are strong departures from normality. The data generating process is given in Definition 1.

Definition 1

$$\epsilon_t = H_t^{1/2} v_t$$

$$h_t = \begin{bmatrix} 1 \\ 0.7 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.2 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.6 \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix}.$$

For the distributional assumption on v_t we take the student t distribution, that is $v_t \sim t_\nu$ with density given in (20). In this exercise we take $\nu = 5$. The results are displayed in Table 2.

Concerning the bias, the three estimation procedures perform similarly, perhaps one could notice that SP performs better than QML for most of the parameters. There are, however, substantial differences between the standard deviations. Clearly ML performs best for all the parameters. SP is as expected in between the two other procedures, also for all the parameters. The same holds true for the MSE of $\hat{\theta}$. One can see that a good part of the loss of the inefficient QML (compared to ML) is recaptured by SP. This has implications, for example, for applications to financial data where often a high kurtosis is observed.

5 Empirical Example

We investigate the bivariate series of daily returns to the CAC 40 (Paris stock exchange) and FTSE 100 (London stock exchange) stock indices from 1990 to 2001 (2894 observations). The two time series are plotted in Figures 1 and 2. Summary statistics of the data are given in Table 3. Obviously, there is excess kurtosis in both series, which is also apparent in the nonparametric density estimates

	ML			*QML*			*SP*		
Population	Mean	SD	MSE	Mean	SD	MSE	Mean	SD	MSE
$c_{11} = 1$	1.021	0.209	0.0441	1.043	0.286	0.0841	1.008	0.249	0.0624
$c_{21} = 0.7$	0.725	0.245	0.0609	0.739	0.309	0.0975	0.736	0.308	0.0967
$c_{22} = 1$	1.039	0.218	0.0493	1.060	0.285	0.0853	1.033	0.258	0.0677
$\beta_{11} = 0.5$	0.492	0.081	0.0066	0.483	0.113	0.0131	0.495	0.097	0.0096
$\beta_{22} = 0.1$	0.070	0.269	0.0735	0.060	0.324	0.1066	0.070	0.320	0.1036
$\beta_{33} = 0.6$	0.591	0.062	0.0040	0.581	0.084	0.0075	0.592	0.075	0.0057
$\alpha_{11} = 0.2$	0.201	0.038	0.0014	0.201	0.055	0.0030	0.194	0.043	0.0019
$\alpha_{22} = 0.1$	0.099	0.032	0.0010	0.102	0.043	0.0018	0.097	0.035	0.0012
$\alpha_{33} = 0.2$	0.201	0.035	0.0012	0.208	0.052	0.0028	0.196	0.039	0.0015

ML, QML and SP Monte Carlo results based on 500 replications of the diagonal VEC model defined in Definition 1 with $n = 2000$. The innovation density is a bivariate student density with 5 degrees of freedom. MSE means mean squared error and SD means standard deviation.

Table 2: Monte Carlo results ($n = 2000, \nu = 5$).

in Figures 3 and 4. For the nonparametric density estimation we used the Gaussian kernel and biased cross validation for the bandwidth. The plots show a superimposed kernel density estimate of data points simulated from a normal distribution with the two moments equal to their sample counterparts. Kolmogorov-Smirnov and Chi-square tests easily reject normality. Also, the Jarque Bera normality test clearly rejects. The unconditional correlation between the two return series is 0.71.

	CAC	FTSE
Min	-0.076	-0.042
Mean	0.31E-3	0.29E-3
Max	0.068	0.054
Std Dev.	0.013	0.010
Skewness	0.178	0.019
E. Kurt.	1.983	1.745

Table 3: Descriptive Statistics

We first estimate a bivariate BEKK(1,1) model by QMLE, which gives us the parameter estimates, reported in Table 4, with a likelihood value of -7604.0.

	QML		*SP*	
C_{11}	0.3452	(4.012)	0.3103	(3.606)
C_{21}	0.0316	(0.9902)	0.0251	(0.7865)
C_{22}	0.0706	(2.500)	0.0695	(2.461)
B_{11}	0.8780	(18.12)	0.8919	(18.40)
B_{21}	0.0715	(1.610)	0.0648	(1.459)
B_{12}	-0.0081	(-0.4111)	-0.0067	(-0.3400)
B_{22}	0.9836	(53.33)	0.9830	(53.29)
A_{11}	0.3192	(6.269)	0.2958	(5.809)
A_{21}	-0.0558	(-0.7998)	-0.0599	(-0.8585)
A_{12}	0.0312	(0.9381)	0.0298	(0.8960)
A_{22}	0.1730	(3.981)	0.1651	(3.799)

Table 4: Parameter estimates with QMLE t-statistics in parentheses.

Note that the off-diagonal elements of A_1 and B_1 are not significant. However, a diagonal BEKK(1,1) model is rejected using a likelihood ratio test statistic. From the estimated model we obtain standardized residuals, $\hat{v}_t$, which are not rejected to be white noise. More precisely, the multivariate Portmanteau statistic with ten lags is 54.11, which is smaller than the 95% critical value of a χ^2 distribution with 40 degrees of freedom.

Although $\hat{v}_t$ is not rejected to be a white noise vector, its components do not appear to be independent. For example, the co-kurtosis is given by $1/n \sum_t \hat{v}_{1t}^2 \hat{v}_{2t}^2 = 1.35$, where one would expect a value of one under independence. A formal test of the hypothesis $E[v_{1t}^2 v_{2t}^2] = 1$ is not known to us unless one uses a specific distribution. But we can use the empirical standard deviation of $v_{1t}^2 v_{2t}^2$ which, divided by $\sqrt{n}$, gives an estimate of the standard error. This value is 0.158, so that under the assumption of an asymptotic normal distribution the estimate of 1.35 is significantly different from 1. We take this rejection of independence as support for our semiparametric procedure since the nonparametric multivariate density estimator automatically captures the dependence among the components of v_t. This contrasts the neglect of any dependence by QMLE where by construction the components of v_t are assumed to be independent.

The estimated bivariate density of $\hat{v}_t$ after the first QMLE estimation step is shown in Figure 5. This function $\hat{g}$ is used in the optimization of $\log L$ in (5). For the semiparametric model, we then obtain the parameter estimates, also reported in Table 4, with a likelihood value of -7392.8. The difference in the likelihood values indicates that the semiparametric model provides a much better fit to the data than the Gaussian model. The multivariate Portmanteau statistic with ten lags applied to the standardized residuals is 53.96, which is slightly smaller than for the Gaussian model.

6 Conclusions and outlook

This paper has shown that efficiency gains of semiparametric GARCH models over QMLE carry over to multivariate GARCH models. In general, the efficiency gain is higher for large sample sizes and for a larger deviation from normality (e.g. skewness and leptokurtosis). We suggest a semiparametric estimator that is shown to be efficient if the true distribution is known to be spherical, and provide conditions under which it is also adaptive. Also, we conjecture that the efficiency gains in the multivariate case increase if the innovations are only uncorrelated but not independent. That is, for example, if the co-kurtosis $E[v_{it}^2 v_{jt}^2]$ is different from 1 for some $i \neq j$. Since our semiparametric procedure estimates the joint density nonparametrically, these dependencies should be captured and yield more efficient parameter estimates.

Appendix A: Figures

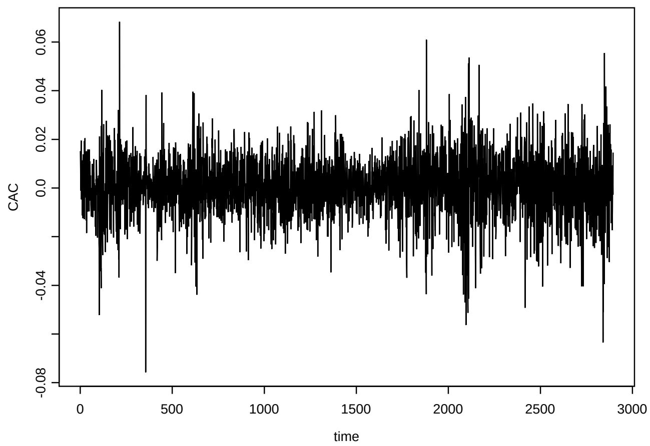

Figure 1: Daily returns for the CAC index from 1/03/1990 to 4/12/2001

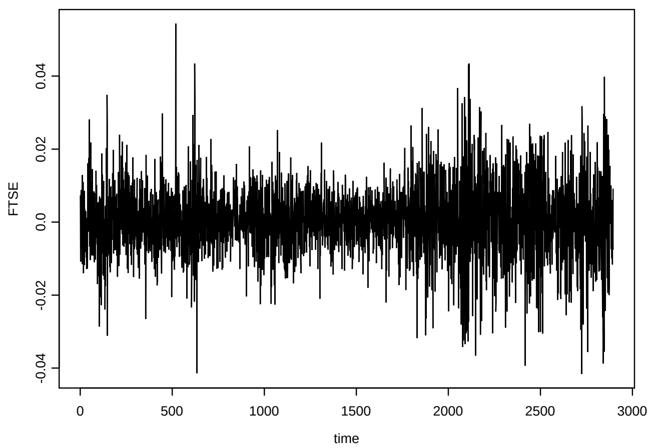

Figure 2: Daily returns for the FTSE index from 1/03/1990 to 4/12/2001

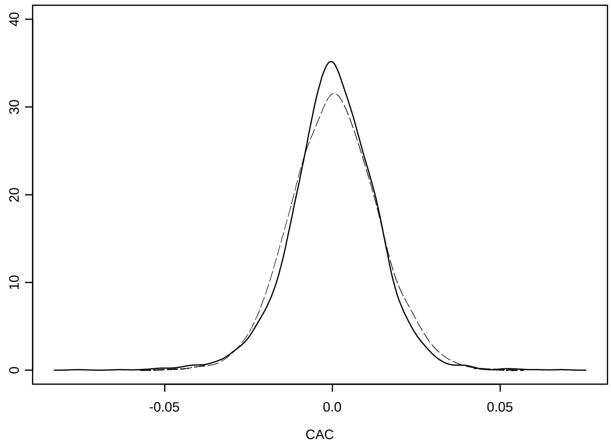

Figure 3: Kernel densities for the daily CAC returns index from 1/03/1990 to 4/12/2001

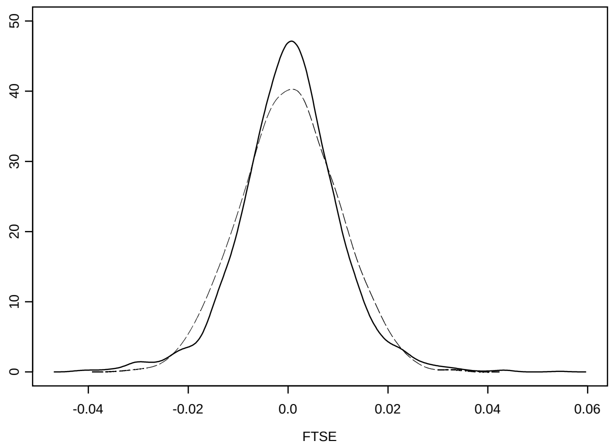

Figure 4: Kernel densities for the daily FTSE returns index from 1/03/1990 to 4/12/2001

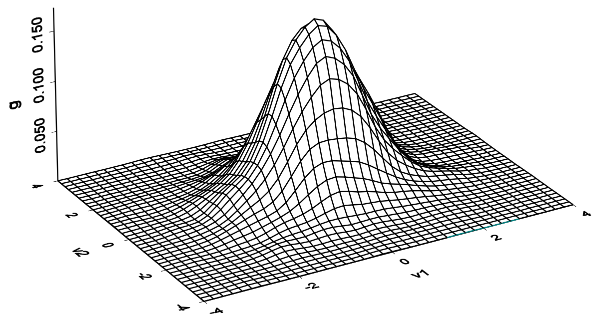

Figure 5: Bivariate kernel density estimate of $\hat{v}_t$ after the QMLE step.

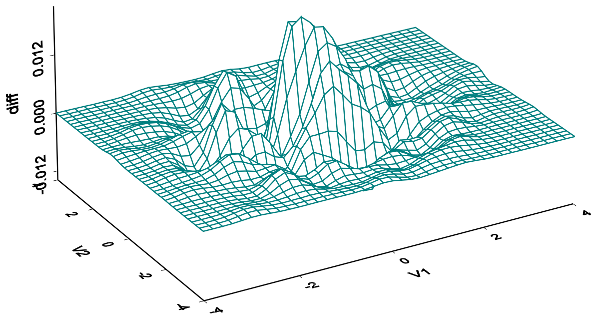

Figure 6: Difference of the bivariate kernel density estimate of $\hat{v}_t$ and a kernel density estimate of n simulated bivariate standard normal r.v.

Appendix B: Index Set Definitions

Define the index

$$k_{ij}^N = i + (j-1)(N - \frac{j}{2}) \tag{23}$$

and the index sets

$$\mathcal{K}_{ij}^N = \begin{cases} \emptyset & N = 1 \\ \{k_{ij}^N \mid j = 1, \ldots, N-1; \quad i = j+1, \ldots, N\} & N \geq 2 \end{cases} \tag{24}$$

and

$$\mathcal{K}_{ii}^N = \{k_{ii}^N \mid i = 1, \ldots, N\} \tag{25}$$

The index k_{ij}^N is the position of the (i,j)-th element of an $(N \times N)$ symmmetric matrix A in the vector vech(A). Remember that vech(A) contains $N^* = N(N+1)/2$ elements. $\mathcal{K}_{ij}^N$ contains all indices of the elements below the diagonal of A in the vector vech(A), this set contains $N(N-1)/2$ elements. The set $\mathcal{K}_{ii}^N$ contains all indices of the N diagonal elements. For example, for $N = 2$, $\mathcal{K}_{ij}^2 = \{2\}$ and $\mathcal{K}_{ii}^2 = \{1,3\}$, and for $N = 3$, $\mathcal{K}_{ij}^3 = \{2,3,5\}$ and $\mathcal{K}_{ii}^3 = \{1,4,6\}$. Note that $\mathcal{K}_{ij}^N \cup \mathcal{K}_{ii}^N = \{1, \ldots, N^*\}$ and $\mathcal{K}_{ij}^N \cap \mathcal{K}_{ii}^N = \emptyset$.

Appendix C: Lemmata

Lemma 1 *For a given matrix $A(m \times m)$, let $Z = I_m \otimes A + A \otimes I_m$, nonsingular. Then*

$$D_m D_m^+ Z^{-1} D_m D_m^+ = D_m D_m^+ Z^{-1} \tag{26}$$

Proof: Indeed

$$\begin{aligned} D_m D_m^+ Z^{-1} D_m D_m^+ Z &= D_m D_m^+ && (27) \\ D_m D_m^+ Z^{-1} Z D_m D_m^+ &= D_m D_m^+ && (28) \\ D_m D_m^+ D_m D_m^+ &= D_m D_m^+ && (29) \end{aligned}$$

by using (70) and the fact that $D_m^+ D_m = I_{m(m+1)/2}$. □

Lemma 2 *For given matrices $A, B(m \times m)$,*

$$D_m'(A \otimes B) D_m D_m^+ = \frac{1}{2} D_m'(A \otimes B + B \otimes A) \tag{30}$$

Proof: Indeed

$$\begin{aligned} D_m'(A \otimes B) D_m D_m^+ &= \frac{1}{2} D_m' \{A \otimes B + C_{mm}(A \otimes B) C_{mm}\} && (31) \\ (D_m D_m^+ \otimes D_m') \text{vec}(A \otimes B) &= \frac{1}{2}(I_{m^2} \otimes D_m' + C_{mm} \otimes D_m' C_{mm}) \text{vec}(A \otimes B) \end{aligned}$$

using first (65) and then vectorizing both sides. But

$$
\begin{aligned}
\frac{1}{2}(I_{m^2} \otimes D_m' + C_{mm} \otimes D_m' C_{mm}) &= \frac{1}{2}(I_{m^2} \otimes D_m' + C_{mm} \otimes D_m') && (32)\\
&= \frac{1}{2}(I_{m^2} + C_{mm}) \otimes D_m' && (33)\\
&= (D_m D_m^+) \otimes D_m', && (34)
\end{aligned}
$$

where (32) uses (68), (34) uses (67). □

Lemma 3 *The matrix $D_N^+ D_N^{+\prime}$ is a $(N^* \times N^*)$ diagonal matrix with 1 at the (i,i)-th position, $i \in \mathcal{K}_{ii}^N$, and 1/2 at the (j,j)-th position, $j \in \mathcal{K}_{ij}^N$, where $\mathcal{K}_{ij}^N$ is defined in (24) and $\mathcal{K}_{ii}^N$ in (25).*

Proof: The statement holds for $D_1^+ D_1^{+\prime} = 1$. Noting that $k_{ii}^{N+1} = k_{i-1,i-1}^N + N + 1$, where k_{ij}^N is defined in (23), and using the recursive equation for $D_{N+1}^+ D_{N+1}^{+\prime}$ in (72), the statement follows by induction. Q.E.D.

Lemma 4 *The matrix $vech(I_N)vech(I_N)'$ is a $(N^* \times N^*)$ matrix with 1 at the (i,j)-th position, $i,j \in \mathcal{K}_{ii}^N$, and 0 elsewhere, where $\mathcal{K}_{ii}^N$ is defined in (25).*

Proof: By definition, $\mathcal{K}_{ii}^N$ contains the positions of the diagonal elements of a $(N \times N)$ matrix A in the vector vech(A). Therefore, $\mathcal{K}_{ii}^N$ contains the positions of the ones in the vector vech(I_N). The matrix vech(I_N)vech$(I_N)'$ then contains ones at pairs of any permutations of these positions, and zeros elsewhere. So there is a total of N^2 ones in the matrix. Q.E.D.

Lemma 5 *For any spherical distribution,*

$$
E\left[\prod_{j=1}^{N} X_j^{\alpha_j}\right] = \begin{cases} 0 & \textit{if one (or more) } \alpha_j \textit{ is odd} \\ K_\alpha \prod_{j=1}^N \frac{\alpha_j!}{(\alpha_j/2)!} & \textit{if all } \alpha_j \textit{ are even} \end{cases}
$$

where $\alpha = \sum_{j=1}^N \alpha_j$ and K_α depends on α only.

Proof: see Box and Hunter (1957).

Appendix D: Proofs

Proof of Proposition 1:

Let us write the likelihood as $L(\theta) = \sum_{t=1}^n l_t(\theta)$ with

$$
l_t = -\frac{1}{2}\log|H_t| + \log g(H_t^{-1/2}\varepsilon_t).
$$

The score vector is given by

$$
\frac{\partial l_t(\theta)}{\partial \theta} = -\frac{1}{2}\frac{\partial \log|H_t|}{\partial \theta} + \frac{\partial \log g(H_t^{-1/2}\varepsilon_t)}{\partial \theta}
$$

where the first term has components

$$\frac{\partial \log |H_t|}{\partial \theta_i} = \mathrm{vec}(H_t^{-1})' \frac{\partial \mathrm{vec}(H_t)}{\partial \theta_i} = \mathrm{Tr}\left(H_t^{-1} \frac{\partial H_t}{\partial \theta_i} \right) = \mathrm{Tr}\left(H_t^{-1/2} \frac{\partial H_t}{\partial \theta_i} H_t^{-1/2} \right)$$

using (78). With the chain rule for matrix differentiation (73), we can write

$$\frac{\partial \log g(H_t^{-1/2}\varepsilon_t)}{\partial \theta_i} = \frac{\partial \log g(x)}{\partial x'} \frac{\partial (H_t^{-1/2}\varepsilon_t)}{\partial \mathrm{vec}(H_t)'} \frac{\partial \mathrm{vec}(H_t)}{\partial \theta_i}$$

Applying (75) and (77) we can further write

$$\frac{\partial (H_t^{-1/2}\varepsilon_t)}{\partial \mathrm{vec}(H_t)'} = -(\varepsilon_t' \otimes I_N)(H_t^{-1/2} \otimes H_t^{-1/2}) \frac{\partial \mathrm{vec}(H_t^{1/2})}{\partial \mathrm{vec}(H_t)'} \tag{35}$$

Then, $\frac{\partial \mathrm{vec}(H_t^{1/2})}{\partial \mathrm{vec}(H_t)'}$ can be appropriately defined by noting that H_t is symmetric and by the definition of the matrix square root (60) $H_t^{1/2}$ is symmetric as well. By (76) we know that in this case

$$\begin{aligned} \frac{\partial \mathrm{vech}(H_t)}{\partial \mathrm{vech}(H_t^{1/2})'} &= D_N^+(H_t^{1/2} \otimes I_N + I_N \otimes H_t^{1/2}) D_N & (36)\\ &= D_N^+ Z D_N & (37) \end{aligned}$$

with $Z = (H_t^{1/2} \otimes I_N + I_N \otimes H_t^{1/2})$, where D_N^+ denotes the generalized inverse (62) of the duplication matrix D_N. If the matrix Z is invertible, a natural definition for the derivative of the matrix square root is

$$\begin{aligned} \frac{\partial \mathrm{vech}(H_t^{1/2})}{\partial \mathrm{vech}(H_t)'} &= (D_N^+ Z D_N)^{-1} & (38)\\ &= D_N^+ Z^{-1} D_N, & (39) \end{aligned}$$

where (39) uses (71). Using (74) we then obtain

$$\begin{aligned} \frac{\partial \mathrm{vec}(H_t^{1/2})}{\partial \mathrm{vec}(H_t)'} &= D_N D_N^+ Z^{-1} D_N D_N^+ & (40)\\ &= D_N D_N^+ Z^{-1} & (41) \end{aligned}$$

by Lemma 1.

Plugging (41) into (35), we obtain

$$\begin{aligned} \frac{\partial (H_t^{-1/2}\varepsilon_t)}{\partial \mathrm{vec}(H_t)'} &= -(\varepsilon_t' \otimes I_N)(H_t^{-1/2} \otimes H_t^{-1/2}) D_N D_N^+ Z^{-1} & (42)\\ &= -(v_t' \otimes I_N)(H_t^{1/2} \otimes I_N)(H_t^{-1/2} \otimes H_t^{-1/2}) Z^{-1} D_N D_N^+ & (43)\\ &= -(v_t' \otimes I_N)(I_N \otimes H_t^{-1/2}) Z^{-1} D_N D_N^+ & (44)\\ &= -(v_t' \otimes I_N)\{Z(I_N \otimes H_t^{1/2})\}^{-1} D_N D_N^+ & (45)\\ &= -(v_t' \otimes I_N)(I_N \otimes H_t + I_{N^2})^{-1} D_N D_N^+ & (46) \end{aligned}$$

where (43) uses (69), (45) uses (61). Thus,

$$\frac{\partial \log g(v_t)}{\partial \theta_i} = -\frac{\partial \log g(v_t)}{\partial v_t'} (v_t' \otimes I_N)(I_N \otimes H_t + I_{N^2})^{-1} D_N D_N^+ \frac{\partial \mathrm{vec}(H_t)}{\partial \theta_i}$$

$$= \quad -\text{vec}\left(\frac{\partial \log g(v_t)}{\partial v_t} v_t'\right)' (I_N \otimes H_t + I_{N^2})^{-1} D_N D_N^+ \frac{\partial \text{vec}(H_t)}{\partial \theta_i}$$

and the score vector can be written as

$$\frac{\partial l_t(\theta)}{\partial \theta} = \frac{\partial \text{vec}(H_t)'}{\partial \theta} \left\{ -\frac{1}{2}\text{vec}(H_t^{-1}) - D_N D_N^+ (I_N \otimes H_t + I_{N^2})^{-1} \text{vec}\left(\frac{\partial \log g(v_t)}{\partial v_t} v_t'\right) \right\}$$

Q.E.D.

Proof of Proposition 2

Under Assumption 3, we have $\text{vec}(\frac{\partial \log g(v_t)}{\partial v_t} v_t') = D_N \text{vech}(\frac{\partial \log g(v_t)}{\partial v_t} v_t')$. Furthermore, if we define $Z = (H_t^{1/2} \otimes I_N + I_N \otimes H_t^{1/2})$ and using Lemma 2 in Appendix C, we can write,

$$\begin{aligned} D_N'(I_N \otimes H_t^{-1/2}) D_N D_N^+ Z^{-1} &= \frac{1}{2} D_N'(I_N \otimes H_t^{-1/2} + H_t^{-1/2} \otimes I_N) Z^{-1} \\ &= \frac{1}{2} D_N'(H_t^{-1/2} \otimes H_t^{-1/2}), \end{aligned}$$

by noting that

$$(I_N \otimes H_t^{-1/2} + H_t^{-1/2} \otimes I_N) = (H_t^{-1/2} \otimes H_t^{-1/2})(H_t^{1/2} \otimes I_N + I_N \otimes H_t^{1/2}).$$

So we have that

$$\begin{aligned} \frac{\partial \log g(v_t)}{\partial \theta_i} &= -\frac{1}{2}\text{vec}(\frac{\partial \log g(v_t)}{\partial v_t} v_t')'(H_t^{-1/2} \otimes H_t^{-1/2}) \frac{\partial \text{vec}(H_t)}{\partial \theta_i} \\ &= -\frac{1}{2}\text{Tr}\left(H_t^{-1/2} \frac{\partial H_t}{\partial \theta_i} H_t^{-1/2} \frac{\partial \log g(v_t)}{\partial v_t} v_t'\right) \end{aligned}$$

using (59). The components of the score are

$$\begin{aligned} \dot{l}_{ti}(\theta) &= -\frac{1}{2}\text{Tr}\left\{ H_t^{-1/2} \frac{\partial H_t}{\partial \theta_i} H_t^{-1/2} \left(I_N + \frac{\partial \log g(v_t)}{\partial v_t} v_t' \right) \right\} \\ &= \text{Tr}(W_{ti} \psi_t) \end{aligned}$$

with $W_{ti} = \frac{1}{2} H_t^{-1/2} \frac{\partial H_t}{\partial \theta_i} H_t^{-1/2}$ and $\psi_t = -\left(I_N + \frac{\partial \log g(v_t)}{\partial v_t} v_t'\right)$. Defining the $(K \times N^2)$ matrix $W_t = (\text{vec}(W_{t1}), \ldots, \text{vec}(W_{tK}))'$, we obtain

$$\dot{l}_t(\theta) = W_t(\theta) D_N \text{vech}(\psi_t).$$

Q.E.D.

Proof of Proposition 3

1. $V_{ml}^{-1} = \text{E}[\dot{l}_t \dot{l}_t'] = \text{E}[W_t \psi_t \psi_t' W_t'] = \text{E}[W_t M_{\psi\psi} W_t']$

2. $$V_{sp}^{-1} = \text{E}[\dot{l}_t \dot{l}_t'] - \text{E}[\dot{l}_t P_t'] - \text{E}[P_t \dot{l}_t'] + \text{E}[P_t P_t'].$$

   The second term is

$$\text{E}[\dot{l}_t P_t'] \quad = \quad \text{E}\left[W_t D_N \psi_t (\psi_t' - F_t' M_{FF}^{-1} M_{F\psi}) D_N' \text{E}(W_t')\right]$$

$$= \quad \mathrm{E}[W_t]D_N\mathrm{E}[\psi_t\psi_t' - \psi_t F_t' M_{FF}^{-1} M_{F\psi}]D_N'\mathrm{E}[W_t']$$
$$= \quad \mathrm{E}[W_t]D_N\left(M_{\psi\psi} - M_{\psi F}M_{FF}^{-1}M_{F\psi}\right)D_N'\mathrm{E}[W_t']$$

Similar calculations show that $\mathrm{E}[\dot{l}_t P_t'] = \mathrm{E}[P_t \dot{l}_t'] = \mathrm{E}[P_t P_t']$, which then gives the stated result.

3. Recall that $V_{qml} = \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}$, and thus $V_{qml}^{-1} = \mathcal{J}\mathcal{I}^{-1}\mathcal{J}$ with
$$\mathcal{J} = -\mathrm{E}\left[\frac{\partial^2 l_t^{qml}}{\partial\theta\partial\theta'}\right], \quad \mathcal{I} = \mathrm{E}\left[\frac{\partial l_t^{qml}}{\partial\theta}\frac{\partial l_t^{qml}}{\partial\theta'}\right],$$
and $\frac{\partial l_t^{qml}}{\partial\theta} = W_t D_N F_t$. We have
$$\mathcal{I} = \mathrm{E}[W_t D_N F_t F_t' D_N' W_t'] = \mathrm{E}[W_t D_N M_{FF} D_N' W_t']$$
and
$$\mathcal{J} = -\int_{\mathbb{R}^N} g(x)\frac{\partial^2 l_t^{qml}}{\partial\theta\partial\theta'}dx = \int_{\mathbb{R}^N}\frac{g(x)}{\partial\theta}\frac{\partial l_t^{qml}}{\partial\theta'}dx \tag{47}$$
$$= \int_{\mathbb{R}^N} g(x)\frac{\log g(x)}{\partial\theta}\frac{\partial l_t^{qml}}{\partial\theta'}dx = \mathrm{E}\left[\frac{\log g(x)}{\partial\theta}\frac{\partial l_t^{qml}}{\partial\theta'}\right] \tag{48}$$
$$= \mathrm{E}\left[W_t D_N \psi_t F_t D_N' W_t'\right] = \mathrm{E}\left[W_t D_N M_{\psi F} D_N' W_t'\right] \tag{49}$$

Q.E.D.

Proof of Proposition 4

1. $M_{\psi F}$: Note first that
$$M_{\psi F} = -\mathrm{E}[\mathrm{vech}(\frac{\partial \log g(x)}{\partial x}x')\mathrm{vech}(xx')'] - \mathrm{vech}(I_N)\mathrm{vech}(I_N)'$$
Writing the first term elementwise, one obtains using integration by parts
$$-\int_{\mathbb{R}^N}\frac{\partial g(x)}{\partial x_i}x_i^3 dx = 3\mathrm{E}[x_i^2] = 3$$
for the (i,i)-th element, $i \in \mathcal{K}_{ii}^N$, and
$$-\int_{\mathbb{R}^N}\frac{\partial g(x)}{\partial x_i}x_i x_j^2 dx = \mathrm{E}[x_j^2] = 1$$
for the (i,j)-th element, $i,j \in \mathcal{K}_{ii}^N$, and $i \neq j$. All other elements are zero, because moments containing odd orders are zero for spherical distributions, see Lemma 5. Thus, Lemma 3 and 4 imply that $M_{\psi F} = 2D_N^+ D_N^{+\prime}$.

2. M_{FF}: Note that
$$M_{FF} = \mathrm{E}[\mathrm{vech}(xx')\mathrm{vech}(xx')'] - \mathrm{vech}(I_N)\mathrm{vech}(I_N)'$$
Using Lemma 5, the (i,i)-th element, $i \in \mathcal{K}_{ii}^N$, of the first term is $\mathrm{E}[x_i^4] = 3c$, and the (i,j)-th element, $i,j \in \mathcal{K}_{ii}^N$, $i \neq j$, is $\mathrm{E}[x_i^2 x_j^2] = c$. All other elements are zero. Together with Lemma 4, this implies that (i,i)-th element, $i \in \mathcal{K}_{ii}^N$ of M_{FF} is $3c-1$ and the (i,j)-th element, $i,j \in \mathcal{K}_{ii}^N$, $i \neq j$, is c.

3. $M_{\psi\psi}$ Note that

$$M_{\psi\psi} = \mathrm{E}\left[\mathrm{vech}\left(\frac{\partial \log g(x)}{\partial x}x'\right)\mathrm{vech}\left(\frac{\partial \log g(x)}{\partial x}x'\right)'\right] - \mathrm{vech}(I_N)\mathrm{vech}(I_N)'$$

A typical element of the first term can be written as

$$\mathrm{E}\left[\frac{\partial \log g(x)}{\partial x_i}\frac{\partial \log g(x)}{\partial x_j}x_k x_l\right] = 4\mathrm{E}\left[f^{-2}(x'x)\left(\frac{\partial f(x'x)}{\partial x'x}\right)^2 x_i x_j x_k x_l\right]$$

because $g(x) = f(x'x)$, which is equal to

$$4\int_{\mathbb{R}^N} f^{-1}(x'x)\left(\frac{\partial f(x'x)}{\partial x'x}\right)^2 x_i x_j x_k x_l dx \tag{50}$$

Now the function $h(x'x) = 4f^{-1}(x'x)\left(\frac{\partial f(x'x)}{\partial x'x}\right)^2$ depends on x only through $x'x$, is positive and integrable by Assumption 4. Thus, it is itself a spherical density up to some scale and (50) is just the fourth order moment structure with respect to h. Therefore, Lemma 5 applies to h and we obtain the same structure as for M_{FF}. That is, for $i \neq j$,

$$\mathrm{E}\left[\left(\frac{\partial \log g(x)}{\partial x_i}\right)^2 x_i^2\right] = 3\mathrm{E}\left[\left(\frac{\partial \log g(x)}{\partial x_i}\right)^2 x_j^2\right]$$

and

$$\mathrm{E}\left[\frac{\partial \log g(x)}{\partial x_i}\frac{\partial \log g(x)}{\partial x_j}x_i x_j\right] = \mathrm{E}\left[\left(\frac{\partial \log g(x)}{\partial x_i}\right)^2 x_j^2\right]. \tag{51}$$

Note that (51) also follows immediately by the symmetry assumption (8). Q.E.D.

Proof of Proposition 5 Adaptivity in the class of spherical distributions is possible if and only if $P_t = 0$. To prove that this occurs only for the multinormal distribution, consider first the case $N = 2$. Then P_t in (11) is a vector with three components. Writing the equation system $P_t = 0$ elementwise, the second equation becomes

$$-\frac{\partial \log g(x)}{\partial x_1}x_2 - \frac{1}{c}x_1 x_2 = 0.$$

Using the symmetry of spherical distributions, this yields

$$\frac{\partial g(x)}{\partial x} = -\frac{1}{c}g(x)x \tag{52}$$

whose unique solution is given by $g(x) = const\,\exp(-\frac{1}{2c}x'x)$, which is the multinormal with covariance matrix cI_N. But since we restricted g to have identity covariance matrix, $c = 1$.

To prove the statement for any dimension N, we have to analyze the structure of the matrix $M_{\psi F}M_{FF}^{-1}$. By Lemma 3, $M_{\psi F}$ is a $(N^* \times N^*)$ diagonal matrix with 2 at the (i,i)-th position, $i \in \mathcal{K}_{ii}^N$, and 1 at the (j,j)-th position, $j \in \mathcal{K}_{ij}^N$.

Lemma 3 and 4 imply that M_{FF} is a $(N^* \times N^*)$ matrix with $3c-1$ at the (i,i)-th position, $i \in \mathcal{K}_{ii}^N$; $c-1$ at the (i,j)-th position, $i,j \in \mathcal{K}_{ii}^N$, $i \neq j$; c at the (i,j)-th position, $i,j \in \mathcal{K}_{ij}^N$;

and zeros elsewhere. Thus, the i-th row and i-th column, $i \in \mathcal{K}_{ij}^N$, of M_{FF} contains a c at the i-th position and zeros elsewhere, and the same row and column of M_{FF}^{-1} contains a $1/c$ at the i-th position and zeros elsewhere. This proves that the $i-th$ element, $i \in \mathcal{K}_{ij}^N$, of the vector $M_{\psi F} M_{FF}^{-1} F_t$ is equal to $v_{ti} v_{tj}/c$. One then obtains the same differential equation (52) with unique solution the N-variate normal distribution. Q.E.D.

Appendix E: Derivatives of the VEC model

For the sake of notational simplicity we take the VEC model with $p = q = 1$

$$h_t = \omega + A\eta_{t-1} + Bh_{t-1} \tag{53}$$

where $\eta_{t-1} = \text{vech}\epsilon_{t-1}\epsilon_{t-1}'$. Deriving with respect to the parameters in ω, A and B we get

$$\begin{aligned} \frac{\partial h_t}{\partial \omega'} &= I_N + B\partial h_{t-1}\partial\omega' && (54)\\ \frac{\partial h_t}{\partial \text{vec}(A)'} &= \eta_{t-1}' \otimes I_N + A\frac{\partial h_{t-1}}{\partial \text{vec}(A)'} && (55)\\ \frac{\partial h_t}{\partial \text{vec}(B)'} &= \eta_{t-1}' \otimes I_N + B\frac{\partial h_{t-1}}{\partial \text{vec}(B)'} && (56) \end{aligned}$$

where (54) is a (3×3) matrix and (55) and (56) are matrices of dimension (3×9) in the bivariate case.

Appendix F: Matrix algebra and calculus

The main part of the following results come from Lütkepohl (1996), abbreviated L hereafter.

1. For matrices A, B, C, D of appropriate dimension, we have

$$\begin{aligned} \text{vec}(ABC) &= (C' \otimes A)\text{vec}(B) && (57)\\ (A \otimes C)(B \otimes D) &= (AB) \otimes (CD) && (58)\\ Tr(ABCD) &= \text{vec}(D')'(C' \otimes A)\text{vec}(B) && (59) \end{aligned}$$

2. Matrix square root: The square root of a symmetric positive definite matrix X is defined as

$$X^{1/2} = \Gamma\Lambda^{1/2}\Gamma' \tag{60}$$

where the columns of Γ contain the eigenvectors of X and $\Lambda^{1/2}$ is diagonal with the positive square roots of the eigenvalues on its diagonal. Note that $X^{1/2}$ is symmetric and positive definite.

3. L 3.5.1 (1), p.27: $X, Y(m \times m)$ nonsingular:

$$(XY)^{-1} = Y^{-1}X^{-1} \tag{61}$$

4. The (Moore-Penrose) generalized inverse of an $(m \times n)$ matrix X can be defined as

$$X^+ = (X'X)^{-1}X' \tag{62}$$

if $X'X$ is nonsingular.

5. The $(mn \times mn)$ commutation matrix C_{mn} is defined by

$$C_{mn}\text{vec}(A) = \text{vec}(A') \tag{63}$$

for every $(m \times n)$ matrix A. Let E_{ij}^{mn} be the $(m \times n)$ matrix with 1 in its ij-th position and zeros elsewhere. Then an explicit expression for C_{mn} is given by

$$C_{mn} = \sum_{i=1}^{m}\sum_{j=1}^{n}(E_{ij}^{mn} \otimes E_{ij}^{mn\prime}). \tag{64}$$

For example, C_{22} is given by

$$C_{22} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

L 9.2.2 (5b), p.117: $A(m \times n), B(p \times q)$:

$$B \otimes A = C_{pm}(A \otimes B)C_{nq} \tag{65}$$

6. The $(n^2 \times n(n+1)/2)$ duplication matrix D_n is defined so that

$$D_n\text{vech}(A) = \text{vec}(A) \tag{66}$$

for every symmetric matrix A of order n. For example, D_2 is given by

$$D_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

An explicit expression is given by

$$D_n = \sum_{j=1}^{n}(\sum_{i>j}^{n} \text{vec}(E_{ij}^{nn} + E_{ji}^{nn})\text{vech}(E_{ij}^{nn})' + \text{vec}(E_{jj}^{nn})\text{vech}(E_{jj}^{nn})').$$

7. L 9.5.2 (1), p.123: The matrix DD^+ is linked to the commutation matrix by

$$D_mD_m^+ = (I_{m^2} + C_{mm})/2 \tag{67}$$

8. L 9.5.2 (2), p.123:

$$C_{mm}D_m = D_m \tag{68}$$

9. L 9.5.4 (1), p.124: $A(m \times m)$:

$$D_m D_m^+ (A \otimes A) = (A \otimes A) D_m D_m^+ \tag{69}$$

10. Theorem 3.11 (iii) Magnus (1988, p.49): $A, B(m \times m)$:

$$\begin{aligned} D_m D_m^+ (A \otimes B + B \otimes A) D_m D_m^+ &= D_m D_m^+ (A \otimes B + B \otimes A) \\ &= (A \otimes B + B \otimes A) D_m D_m^+ \end{aligned} \tag{70}$$

11. L 9.5.4 (8c), p. 125: $A(m \times m)$:

$$\left(D_m^+ \left(I_m \otimes A + A \otimes I_m\right) D_m\right)^{-1} = D_m^+ \left(I_m \otimes A + A \otimes I_m\right)^{-1} D_m \tag{71}$$

12. L, p. 125:

$$D_{m+1}^+ D_{m+1}^{+\prime} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} I_m & 0 \\ 0 & 0 & D_m^+ D_m^{+\prime} \end{bmatrix} \tag{72}$$

13. Chain rule for matrix differentiation, L 10.7(2), p.203: $X(m \times n)$, $Y(X)(p, \times q)$, $Z(Y)(r \times s)$:

$$\frac{\partial \mathrm{vec}(Z(Y(X)))}{\partial \mathrm{vec}(X)'} = \frac{\partial \mathrm{vec}(Z(Y))}{\partial \mathrm{vec}(Y)'} \frac{\partial \mathrm{vec}(Y(X))}{\partial \mathrm{vec}(X)'} \tag{73}$$

14. Magnus (1988, p.129): $X(m \times m)$ symmetric, $Y(X)$ symmetric matrix function: Using (66), the differential of $\mathrm{vec}(Y)$ can be written as

$$\begin{aligned} \mathrm{dvec}(Y) &= D_m \frac{\partial \mathrm{vech}(Y)}{\partial \mathrm{vech}(X)'} \mathrm{dvech}(X) \\ &= D_m \frac{\partial \mathrm{vech}(Y)}{\partial \mathrm{vech}(X)'} D_m^+ \mathrm{dvec}(X) \end{aligned} \tag{74}$$

15. L 10.4 (3), p.183: $X(m \times n)$, $A(p \times m)$, $B(n \times q)$:

$$\frac{\partial \mathrm{vec}(AXB)}{\partial \mathrm{vec}(X)'} = B' \otimes A \tag{75}$$

16. L 10.5.3 (2), p. 194: $X, A(m \times m)$ symmetric:

$$\frac{\partial \mathrm{vech}(XAX)}{\partial \mathrm{vech}(X)'} = D_m^+ (XA \otimes I_m + I_m \otimes XA) D_m \tag{76}$$

17. L 10.6 (1), p.198: $X(m \times m)$ nonsigular:

$$\frac{\partial \mathrm{vec}(X^{-1})}{\partial \mathrm{vec}(X)'} = -X'^{-1} \otimes X^{-1} \tag{77}$$

18. L 10.3.3, p.182: $X(m \times m)$, $|X| > 0$:

$$\frac{\partial \log |X|}{\partial X} = (X')^{-1} \tag{78}$$

References

[1] Bauwens, L., Laurent, S., Rombouts, J.V.K., 2002, A Review of Multivariate GARCH Models with Application to Financial Data, mimeo, UCL.

[2] Bickel, P.J., 1982, On adaptive Estimation, Annals of Statistics, 10, 647-671.

[3] Bickel, P.J., Klaassen, C.A.J., Ritov, Y., Wellner, J.A., 1993, Efficient and Adaptive Statistical Inference for Semiparametric Models. John Hopkins University Press, Baltimore, MD.

[4] Bollerslev, T., Wooldridge, J., 1992, Quasi Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances, Econometric Reviews, 11, 143-72.

[5] Box, G.E.P., Hunter, J.S., 1957, Multi-factor experimental designs for exploring response surfaces, Annals of Mathematical Statistics, 28, 195–241.

[6] Drost, F.C., Klaassen, C.A.J., Werker, B.J.M., 1997, Adaptive Estimation in Time Series Models, Annals of Statistics, 25, 786-817.

[7] Drost, C.D., Klaassen, C.A.J, 1997, Efficient Estimation in Semiparametric GARCH Models, Journal of Econometrics, 81, 193-221.

[8] Comte, F., Lieberman, O., 2000, Asymptotic Theory for Multivariate GARCH Processes, working paper, Unversities of Paris 6 and 7.

[9] Engle, R. and Kroner, K., 1995, Multivariate Simultaneous GARCH, Econometric Theory, 11, 122-150.

[10] Engle, R., Gonzalez-Rivera, G., 1991, Semiparametric ARCH Models, Journal of Business and Economic Statistics, 9/4, 345-359.

[11] Gonzalez-Rivera, G., 1997, A Note on Adaptation in ARCH models, Econometric Reviews, 16, 55-68.

[12] Gonzalez-Rivera, G., Drost, C.D., 1999, Efficiency Comparisons of Maximum-Likelihood-based Estimators in GARCH Models, Journal of Econometrics, 93, 93-111.

[13] Jeantheau, T., 1998, Strong consistency of estimators for multivariate ARCH models, Econometric Theory 14, 70-86.

[14] Jensen, D.R., 1985, Multivariate Distributions, Encyclopedia of Statistical Sciences, 6, (eds S., Kotz, N.L., Johnson and C.B. Read), Wiley, 43-55.

[15] Linton, O., 1993, Adaptive Estimation of ARCH models, Econometric Theory, 9, 539-569.

[16] Lütkepohl, H., 1996, *Handbook of Matrices*, New York: Wiley.

[17] Magnus, J.R., 1988, *Linear Structures*, London: Griffin.

[18] Newey, W.K., 1990, Semiparametric Efficiency Bounds, Journal of Applied Econometrics, 5, 99-135.

[19] Scott, D., 1992, *Multivariate Density estimation: theory, practice, and visualisation*, New York: Wiley.

[20] Silverman, B.W., 1986, *Density estimation for statistics and data analysis*, Chapman and Hall: London.